PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/47, C12N 15/10

A1

(11) International Publication Number: **WO 99/36521**

(43) International Publication Date: 22 July 1999 (22.07.99)

(21) International Application Number: PCT/CN98/00010

(22) International Filing Date: 19 January 1998 (19.01.98)

(71) Applicant *(for all designated States except US)*: SHANGHAI SECOND MEDICAL UNIVERSITY [CN/CN]; 197 Rui-Jin Road II, Shanghai 200025 (CN).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MAO, Mao [CN/CN]; 50 Gu Mei Apartment #3, Room 301, 360 Pingyang Road, Shanghai 200233 (CN). FU, Gang [CN/CN]; Room 104, 43 Long Shan Xin Cun, Shanghai 200030 (CN). ZHANG, Qing-hua [CN/CN]; Ruijin Hospital, 21st Building, Room 101, 197 Ruijin Road II, Shanghai 200025 (CN). ZHOU, Juan [CN/CN]; Room 303, 11,290 Pudong Avenue, Shanghai 200120 (CN).

(74) Agent: CHINA PATENT AGENT (H.K.) LTD.; 22/F., Great Eagle Centre, 23 Harbour Road, Wanchai, Hong-Kong (CN).

(81) Designated States: CA, CN, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: A GENE HOMOLOGOUS TO FOX SPERM ACROSOMAL PROTEIN FSA-1 (CBCALD05)

(57) Abstract

CBCALD05 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing CBCALD05 polypeptides and polynucleotides in the design of protocols for the treatment of infertility and diseases related to fertility, among others, and diagnostic assays for such conditions.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

A Gene Homologous to Fox Sperm Acrosomal Protein FSA-1 (CBCALD05)

FIELD OF INVENTION

This invention relates to newly identified polynucleotides, polypeptides encoded by them and to the use of such polynucleotides and polypeptides, and to their production. More particularly, the polynucleotides and polypeptides of the present invention relate to the gene family containing acrosomal protein, hereinafter referred to as CBCALD05. The invention also relates to inhibiting or activating the action of such polynucleotides and polypeptides.

BACKGROUND OF THE INVENTION

Te fox FSA-1 is a membrane protein, located on the acrosomal membrane, and it has several partners in many mammals, and it may involve in fsperm activity and/or fertilization. The novel human candidate for this protein has little homology to the known human sperm acrosomal proteins. This indicates that the Gene family containing acrosomal protein has an established, proven history as therapeutic targets. Clearly there is a need for identification and characterization of further members of the Gene family containing acrosomal protein which can play a role in preventing, ameliorating or correcting dysfunctions or diseases, including, but not limited to, infertility and diseases related to fertility.

SUMMARY OF THE INVENTION

In one aspect, the invention relates to CBCALD05 polypeptides and recombinant materials and methods for their production. Another aspect of the invention relates to methods for using such CBCALD05 polypeptides and polynucleotides. Such uses include the treatment of infertility and diseases related to fertility, among others. In still another aspect, the invention relates to methods to identify agonists and antagonists using the materials provided by the invention, and treating conditions associated with CBCALD05 imbalance with the identified compounds. Yet another aspect of the invention relates to diagnostic assays for detecting diseases associated with inappropriate CBCALD05 activity or levels.

DESCRIPTION OF THE INVENTION

Definitions

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"CBCALD05" refers, among others, generally to a polypeptide having the amino acid sequence set forth in SEQ ID NO:2 or an allelic variant thereof.

"CBCALD05 activity or CBCALD05 polypeptide activity" or "biological activity of the CBCALD05 or CBCALD05 polypeptide" refers to the metabolic or physiologic function of said CBCALD05 including similar activities or improved activities or these activities with decreased undesirable side-effects. Also included are antigenic and immunogenic activities of said CBCALD05.

"CBCALD05 gene" refers to a polynucleotide having the nucleotide sequence set forth in SEQ ID NO:1 or allelic variants thereof and/or their complements.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications has been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded

amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol* (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Posttranslational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62.

"Variant" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions,

deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity" is a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. See, e.g.: (COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, 1988; BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, 1993; COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, 1987; and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCS program package (Devereux, J., *et al.*, *Nucleic Acids Research* (1984) 12(1):387), BLASTP, BLASTN, FASTA (Atschul, S.F. *et al.*, *J Molec Biol* (1990) 215:403).

As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence of SEQ ID NO: 1 is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence of SEQ ID NO: 1. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5 or 3 terminal positions of the reference nucleotide sequence or anywhere between those terminal positions,

interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference amino acid sequence of SEQ ID NO:2 is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of SEQ ID NO: 2. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

Polypeptides of the Invention

In one aspect, the present invention relates to CBCALD05 polypeptides (or CBCALD05 proteins). The CBCALD05 polypeptides include the polypeptide of SEQ ID NO:2; as well as polypeptides comprising the amino acid sequence of SEQ ID NO: 2; and polypeptides comprising the amino acid sequence which have at least 80% identity to that of SEQ ID NO:2 over its entire length, and still more preferably at least 90% identity, and even still more preferably at least 95% identity to SEQ ID NO: 2. Furthermore, those with at least 97-99% are highly preferred. Also included within CBCALD05 polypeptides are polypeptides having the amino acid sequence which have at least 80% identity to the polypeptide having the amino acid sequence of SEQ ID NO:2 over its entire length, and still more preferably at least 90% identity, and still more preferably at least 95% identity to SEQ ID NO:2. Furthermore, those with at least 97-99% are highly preferred. Preferably CBCALD05 polypeptide exhibit at least one biological activity of CBCALD05.

The CBCALD05 polypeptides may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

Fragments of the CBCALD05 polypeptides are also included in the invention. A fragment is a polypeptide having an amino acid sequence that entirely is the same as part, but not all, of the amino acid

sequence of the aforementioned CBCALD05 polypeptides. As with CBCALD05 polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, and 101 to the end of CBCALD05 polypeptide. In this context "about" includes the particularly recited ranges larger or smaller by several, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

Preferred fragments include, for example, truncation polypeptides having the amino acid sequence of CBCALD05 polypeptides, except for deletion of a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Also preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Other preferred fragments are biologically active fragments. Biologically active fragments are those that mediate CBCALD05 activity, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those that are antigenic or immunogenic in an animal, especially in a human.

Preferably, all of these polypeptide fragments retain the biological activity of the CBCALD05, including antigenic activity. Variants of the defined sequence and fragments also form part of the present invention. Preferred variants are those that vary from the referents by conservative amino acid substitutions – i.e., those that substitute a residue with another of like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination.

The CBCALD05 polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

Polynucleotides of the Invention

Another aspect of the invention relates to CBCALD05 polynucleotides. CBCALD05 polynucleotides include isolated polynucleotides which encode the CBCALD05 polypeptides and fragments, and polynucleotides closely related thereto. More specifically, CBCALD05 polynucleotide of the invention

include a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:1 encoding a CBCALD05 polypeptide of SEQ ID NO: 2, and polynucleotide having the particular sequence of SEQ ID NO:1. CBCALD05 polynucleotides further include a polynucleotide comprising a nucleotide sequence that has at least 80% identity over its entire length to a nucleotide sequence encoding the CBCALD05 polypeptide of SEQ ID NO:2, and a polynucleotide comprising a nucleotide sequence that is at least 80% identical to of SEQ ID NO:1 over its entire length. In this regard, polynucleotides at least 90% identical are particularly preferred, and those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred and those with at least 98-99% are most highly preferred, with at least 99% being the most preferred. Also included under CBCALD05 polynucleotides are a nucleotide sequence which has sufficient identity to a nucleotide sequence contained in SEQ ID NO:1 to hybridize under conditions useable for amplification or for use as a probe or marker. The invention also provides polynucleotides which are complementary to such CBCALD05 polynucleotides.

CBCALD05 of the invention is structurally related to other proteins of the Gene family containing acrosomal protein, as shown by the results of sequencing the cDNA of Table 1 (SEQ ID NO:1) encoding human CBCALD05. The cDNA sequence of SEQ ID NO:1 contains an open reading frame (nucleotide number 11 to 879) encoding a polypeptide of 293 amino acids of SEQ ID NO:2. The amino acid sequence of Table 2 (SEQ ID NO:2) has about 82.2% identity (using FASTA) in 236 amino acid residues with fox acrosomal protein FSA-1 (S Beaton et al.,Reprod. Fertil. Dev. 6:761-770,1994). The nucleotide sequence of Table 1 (SEQ ID NO:1) has about 88.5% identity (using FASTA) in 574 nucleotide residues with fox sperm acrosoaml protein FSA-1 (S Beaton et al.,Reprod. Fertil. Dev. 6:761-770,1994). Thus, CBCALD05 polypeptides and polynucleotides of the present invention are expected to have, inter alia, similar biological functions/properties to their homologous polypeptides and polynucleotides, and their utility is obvious to anyone skilled in the art.

Table 1[a]

```
   1  GACTCCCAAG ATGGCGGACC TACTGGGCTC CATCCTGAGC TCCATGGAGA

  51  AGCCACCCAG CCTCGGTGAC CAGGAGACTC GGCGCAAGGC CCGAGAACAG

 101  GCCGCCCGCC TGAAGAAACT ACAAGAGCAA GAGAAACAAC AGAAAGTGGA

 151  GTTTCGTAAA AGGATGGAGA AGGAGGTGTC AGATTTCATT CAAGACAGTG

 201  GGCAGATCAA GAAAAAGTTT CAGCCAATGA ACAAGATCGA GAGGAGCATA

 251  CTACATGATG TGGTGGAAGT GGCTGGCCTG ACATCCTTCT CCTTTGGGGA
```

```
301  AGATGATGAC TGTCGCTATG TCATGATCTT CAAAAAGGAG TTTGCACCCT

351  CAGATGAAGA GCTAGACTCT TACCGTCGTG GAGAGGAATG GGACCCCCAG

401  AAGGCTGAGG AGAAGCGGAA GCTGAAGGAG CTGGCCCAGA GGCAAGAGGA

451  GGAGGCAGCC CAGCAGGGGC CTGTGGTGGT GAGCCCTGCC AGCGACTACA

501  AGGACAAGTA CAGCCACCTC ATCGGCAAGG GAGCAGCCAA AGACGCAGCC

551  CACATGCTAC AGGCCAATAA GACCTACGGC TGTGTGCCCG TGGCCAATAA

601  GAGGGACACA CGCTCCATTG AAGAGGCTAT GAATGAGATC AGAGCCAAGA

651  AGCGTCTGCG GCAGAGTGGG GAAGAGTTGC CGCCAACCTC TAGGCGCCCC

701  GCCCAGCTCC CTTTGACCCC TGGGGCAGGG CAGGGGGCAG GGAGAGACAA

751  GGCTGCTGCT ATTAGAGCCC ATCCTGGAGC CCCACCTCTG AACCACCTCC

801  TACCAGCTGT CCCTCAGGCT GGGGGAAAAC AGGTGTTTGA TTTGTCACCG

851  TTGGAGCTTG GATATGTGCG TGGCATGTGT GTGTGTGTGT GAGAGTGTGA

901  ATGCACAGGT GGGTATTTAA TCTGTATTAT TCCCCGTTCT TGGAATTTTC

951  TTCCCCATGG GGCTGGGGTA CTTTACATTC AATAAATACT GTTTAACCCA

1001 AAAAAAAAAA AAAAAAAAAA AAAAAAA
```

[a] A nucleotide sequence of a human CBCALD05 (SEQ ID NO: 1).

Table 2[b]

```
  1  MADLLGSILS SMEKPPSLGD QETRRKAREQ AARLKKLQEQ EKQQKVEFRK

 51  RMEKEVSDFI QDSGQIKKKF QPMNKIERSI LHDVVEVAGL TSFSFGEDDD

101  CRYVMIFKKE FAPSDEELDS YRRGEEWDPQ KAEEKRKLKE LAQRQEEEAA

151  QQGPVVVSPA SDYKDKYSHL IGKGAAKDAA HMLQANKTYG CVPVANKRDT
```

```
201  RSIEEAMNEI RAKKRLRQSG EELPPTSRRP AQLPLTPGAG QGAGRDKAAA

251  IRAHPGAPPL NHLLPAVPQA GGKQVFDLSP LELGYVRGMC VCV
```

[b] An amino acid sequence of a human CBCALD05 (SEQ ID NO: 2).

One polynucleotide of the present invention encoding CBCALD05 may be obtained using standard cloning and screening, from a cDNA library derived from mRNA in cells of human cord blood using the expressed sequence tag (EST) analysis (Adams, M.D., *et al. Science* (1991) 252:1651-1656; Adams, M.D. *et al., Nature*, (1992) *355*:632-634; Adams, M.D., *et al., Nature* (1995) 377 Supp:3-174). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

The nucleotide sequence encoding CBCALD05 polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in Table 1 (nucleotide number 11 to 879 of SEQ ID NO:1), or it may be a sequence, which as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2.

When the polynucleotides of the invention are used for the recombinant production of CBCALD05 polypeptide, the polynucleotide may include the coding sequence for the mature polypeptide or a fragment thereof, by itself; the coding sequence for the mature polypeptide or fragment in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al., Proc Natl Acad Sci USA* (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further preferred embodiments are polynucleotides encoding CBCALD05 variants comprise the amino acid sequence CBCALD05 polypeptide of Table 2 (SEQ ID NO:2) in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acid residues are substituted, deleted or added, in any combination.

The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent

conditions" means hybridization will occur only if there is at least 80%, and preferably at least 90%, and more preferably at least 95%, yet even more preferably 97-99% identity between the sequences.

Polynucleotides of the invention, which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1 or a fragment thereof, may be used as hybridization probes for cDNA and genomic DNA, to isolate full-length cDNAs and genomic clones encoding CBCALD05 polypeptide and to isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than human) that have a high sequence similarity to the CBCALD05 gene. Such hybridization techniques are known to those of skill in the art. Typically these nucleotide sequences are 80% identical, preferably 90% identical, more preferably 95% identical to that of the referent. The probes generally will comprise at least 15 nucleotides. Preferably, such probes will have at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will range between 30 and 50 nucleotides.

In one embodiment, to obtain a polynucleotide encoding CBCALD05 polypeptide, including homologs and orthologs from species other than human, comprises the steps of screening an appropriate library under stingent hybridization conditions with a labeled probe having the SEQ ID NO: 1 or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Thus in another aspect, CBCALD05 polynucleotides of the present invention further include a nucleotide sequence comprising a nucleotide sequence that hybridize under stringent condition to a nucleotide sequence having SEQ ID NO: 1 or a fragment thereof. Also included with CBCALD05 polypeptides are polypeptide comprising amino acid sequence encoded by nucleotide sequence obtained by the above hybridization condition. Such hybridization techniques are well known to those of skill in the art. Stringent hybridization conditions are as defined above or, alternatively, conditions under overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments and diagnostics to animal and human disease.

Vectors, Host Cells, Expression

The present invention also relates to vectors which comprise a polynucleotide or polynucleotides of the present invention, and host cells which are genetically engineered with vectors of the invention and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into

host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY* (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989) such as calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used. Such systems include, among others, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL* (*supra*).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the desired polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If the CBCALD05 polypeptide is to be expressed for use in screening assays, generally, it is preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If CBCALD05 polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide; if produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

CBCALD05 polypeptides can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high

performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention also relates to the use of CBCALD05 polynucleotides for use as diagnostic reagents. Detection of a mutated form of CBCALD05 gene associated with a dysfunction will provide a diagnostic tool that can add to or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of CBCALD05. Individuals carrying mutations in the CBCALD05 gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled CBCALD05 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers *et al.*, *Science* (1985) 230:1242. Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method. See Cotton *et al.*, *Proc Natl Acad Sci USA* (1985) 85: 4397-4401. In another embodiment, an array of oligonucleotides probes comprising CBCALD05 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability. (See for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to infertility and diseases related to fertility through detection of mutation in the CBCALD05 gene by the methods described.

In addition, infertility and diseases related to fertility, can be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of CBCALD05 polypeptide or CBCALD05 mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as an CBCALD05

polypeptide, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagonostic kit for a disease or suspectability to a disease, particularly infertility and diseases related to fertility, which comprises:

(a) a CBCALD05 polynucleotide, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment thereof ;

(b) a nucleotide sequence complementary to that of (a);

(c) a CBCALD05 polypeptide, preferably the polypeptide of SEQ ID NO: 2, or a fragment thereof; or

(d) an antibody to a CBCALD05 polypeptide, preferably to the polypeptide of SEQ ID NO: 2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

Chromosome Assays

The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

Antibodies

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them can also be used as immunogens to produce antibodies immunospecific for the CBCALD05 polypeptides. The term "immunospecific" means that the antibodies have substantiall greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against the CBCALD05 polypeptides can be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler,

G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, MONOCLONAL ANTIBODIES AND CANCER THERAPY, pp. 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against CBCALD05 polypeptides may also be employed to treat infertility and diseases related to fertility, among others.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with CBCALD05 polypeptide, or a fragment thereof, adequate to produce antibody and/or T cell immune response to protect said animal from infertility and diseases related to fertility, among others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering CBCALD05 polypeptide via a vector directing expression of CBCALD05 polynucleotide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

Further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a CBCALD05 polypeptide wherein the composition comprises a CBCALD05 polypeptide or CBCALD05 gene. The vaccine formulation may further comprise a suitable carrier. Since CBCALD05 polypeptide may be broken down in the stomach, it is preferably administered parenterally (including subcutaneous, intramuscular, intravenous, intradermal etc. injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation instonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Screening Assays

The CBCALD05 polypeptide of the present invention may be employed in a screening process for compounds which activate (agonists) or inhibit activation of (antagonists, or otherwise called inhibitors) the CBCALD05 polypeptide of the present invention. Thus, polypeptides of the invention may also be used to assess identify agonist or antagonists from, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These agonists or antagonists may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide of the present invention; or may be structural or functional mimetics of the polypeptide of the present invention. See Coligan *et al.*, *Current Protocols in Immunology* 1(2):Chapter 5 (1991).

CBCALD05 polypeptides are responsible for many biological functions, including many pathologies. Accordingly, it is desirous to find compounds and drugs which stimulate CBCALD05 polypeptide on the one hand and which can inhibit the function of CBCALD05 polypeptide on the other hand. In general, agonists are employed for therapeutic and prophylactic purposes for such conditions as infertility and diseases related to fertility. Antagonists may be employed for a variety of therapeutic and prophylactic purposes for such conditions as infertility and diseases related to fertility.

In general, such screening procedures may involve using appropriate cells which express the CBCALD05 polypeptide or respond to CBCALD05 polypeptide of the present invention. Such cells include cells from mammals, yeast, *Drosophila* or *E. coli*. Cells which express the CBCALD05 polypeptide (or cell membrane containing the expressed polypeptide) or respond to CBCALD05 polypeptide are then contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The ability of the cells which were contacted with the candidate compounds is compared with the same cells which were not contacted for CBCALD05 activity.

The assays may simply test binding of a candidate compound wherein adherence to the cells bearing the CBCALD05 polypeptide is detected by means of a label directly or indirectly associated with the candidate compound or in an assay involving competition with a labeled competitor. Further, these assays may test whether the candidate compound results in a signal generated by activation of the CBCALD05 polypeptide, using detection systems appropriate to the cells bearing the CBCALD05 polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed.

Further, the assays may simply comprise the steps of mixing a candidate compound with a solution containing a CBCALD05 polypeptide to form a mixture, measuring CBCALD05 activity in the mixture, and comparing the CBCALD05 activity of the mixture to a standard.

The CBCALD05 cDNA, protein and antibodies to the protein may also be used to configure assays for detecting the effect of added compounds on the production of CBCALD05 mRNA and protein in cells. For example, an ELISA may be constructed for measuring secreted or cell associated levels of CBCALD05 protein using monoclonal and polyclonal antibodies by standard methods known in the art, and this can be used to discover agents which may inhibit or enhance the production of CBCALD05 (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The CBCALD05 protein may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the CBCALD05 is labeled with a radioactive isotope (eg 125I), chemically modified (eg biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. In addition to being used for purification and cloning of the receptor, these binding assays can be used to identify agonists and antagonists of CBCALD05 which compete with the binding of CBCALD05 to its receptors, if any. Standard methods for conducting screening assays are well understood in the art.

Examples of potential CBCALD05 polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the CBCALD05 polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypetide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Thus in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for CBCALD05 polypeptides; or compounds which decrease or enhance the production of CBCALD05 polypeptides, which comprises:

(a) a CBCALD05 polypeptide, preferably that of SEQ ID NO:2;

(b) a recombinant cell expressing a CBCALD05 polypeptide, preferably that of SEQ ID NO:2;

(c) a cell membrane expressing a CBCALD05 polypeptide; preferably that of SEQ ID NO: 2; or

(d) antibody to a CBCALD05 polypeptide, preferably that of SEQ ID NO: 2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

Prophylactic and Therapeutic Methods

This invention provides methods of treating abnormal conditions such as, infertility and diseases related to fertility, related to both an excess of and insufficient amounts of CBCALD05 polypeptide activity.

If the activity of CBCALD05 polypeptide is in excess, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as hereinabove described along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the CBCALD05 polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of CBCALD05 polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous CBCALD05 polypeptide may be administered. Typical embodiments of such competitors comprise fragments of the CBCALD05 polypeptide.

In still another approach, expression of the gene encoding endogenous CBCALD05 polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered. See, for example, O'Connor, *J Neurochem* (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988). Alternatively, oligonucleotides which form triple helices with the gene can be supplied. See, for example, Lee *et al.*, *Nucleic Acids Res* (1979) 6:3073; Cooney *et al.*, *Science* (1988) 241:456; Dervan *et al.*, *Science* (1991) 251:1360. These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*.

For treating abnormal conditions related to an under-expression of CBCALD05 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates CBCALD05 polypeptide, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of CBCALD05 by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For overview of gene therapy, see Chapter 20, *Gene Therapy and other Molecular Genetic-based Therapeutic Approaches*, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of CBCALD05 polypeptides in combination with a suitable pharmaceutical carrier.

Formulation and Administration

Peptides, such as the soluble form of CBCALD05 polypeptides, and agonists and antagonist peptides or small molecules, may be formulated in combination with a suitable pharmaceutical carrier. Such formulations comprise a therapeutically effective amount of the polypeptide or compound, and a pharmaceutically acceptable carrier or excipient. Such carriers include but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. Formulation should suit the mode of administration, and is well within the skill of the art. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

Preferred forms of systemic administration of the pharmaceutical compositions include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if properly formulated in enteric or encapsulated formulations, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels and the like.

The dosage range required depends on the choice of peptide, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 μg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: SHANGHAI SECOND MEDICAL UNIVERSITY

(ii) TITLE OF THE INVENTION: A GENE HOMOLOGOUS TO FOX SPERM ACROSOMAL PROTEIN FSA-1 (CBCALD05)

(iii) NUMBER OF SEQUENCES: 2

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: RATNER & PRESTIA
(B) STREET: P.O. BOX 980
(C) CITY: VALLEY FORGE
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19482

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: TO BE ASSIGNED
(B) FILING DATE:
(C) CLASSIFICATION: UNKNOWN

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: PRESTIA, PAUL F
(B) REGISTRATION NUMBER: 23,031
(C) REFERENCE/DOCKET NUMBER: GP-70355

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 610-407-0700
(B) TELEFAX: 610-407-0701
(C) TELEX: 846169

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1027 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GACTCCCAAG ATGGCGGACC TACTGGGCTC CATCCTGAGC TCCATGGAGA AGCCACCCAG 60
CCTCGGTGAC CAGGAGACTC GGCGCAAGGC CCGAGAACAG GCCGCCCGCC TGAAGAAACT 120
ACAAGAGCAA GAGAAACAAC AGAAAGTGGA GTTTCGTAAA AGGATGGAGA AGGAGGTGTC 180
AGATTTCATT CAAGACAGTG GGCAGATCAA GAAAAAGTTT CAGCCAATGA ACAAGATCGA 240
GAGGAGCATA CTACATGATG TGGTGGAAGT GGCTGGCCTG ACATCCTTCT CCTTTGGGGA 300
AGATGATGAC TGTCGCTATG TCATGATCTT CAAAAAGGAG TTTGCACCCT CAGATGAAGA 360
GCTAGACTCT TACCGTCGTG GAGAGGAATG GGACCCCCAG AAGGCTGAGG AGAAGCGGAA 420
GCTGAAGGAG CTGGCCCAGA GGCAAGAGGA GGAGGCAGCC CAGCAGGGGC CTGTGGTGGT 480
GAGCCCTGCC AGCGACTACA AGGACAAGTA CAGCCACCTC ATCGGCAAGG GAGCAGCCAA 540
AGACGCAGCC CACATGCTAC AGGCCAATAA GACCTACGGC TGTGTGCCCG TGGCCAATAA 600
GAGGGACACA CGCTCCATTG AAGAGGCTAT GAATGAGATC AGAGCCAAGA AGCGTCTGCG 660
GCAGAGTGGG GAAGAGTTGC CGCCAACCTC TAGGCGCCCC GCCCAGCTCC CTTTGACCCC 720
TGGGGCAGGG CAGGGGGCAG GGAGAGACAA GGCTGCTGCT ATTAGAGCCC ATCCTGGAGC 780
CCCACCTCTG AACCACCTCC TACCAGCTGT CCCTCAGGCT GGGGGAAAAC AGGTGTTTGA 840
TTTGTCACCG TTGGAGCTTG GATATGTGCG TGGCATGTGT GTGTGTGTGT GAGAGTGTGA 900
ATGCACAGGT GGGTATTTAA TCTGTATTAT TCCCCGTTCT TGGAATTTTC TTCCCCATGG 960
GGCTGGGGTA CTTTACATTC AATAAATACT GTTTAACCCA AAAAAAAAAA AAAAAAAAAA 1020
AAAAAAA 1027

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Asp Leu Leu Gly Ser Ile Leu Ser Ser Met Glu Lys Pro Pro
 1               5                  10                  15
Ser Leu Gly Asp Gln Glu Thr Arg Arg Lys Ala Arg Glu Gln Ala Ala
            20                  25                  30
Arg Leu Lys Lys Leu Gln Glu Gln Glu Lys Gln Gln Lys Val Glu Phe
        35                  40                  45
Arg Lys Arg Met Glu Lys Glu Val Ser Asp Phe Ile Gln Asp Ser Gly
    50                  55                  60
Gln Ile Lys Lys Lys Phe Gln Pro Met Asn Lys Ile Glu Arg Ser Ile
65                  70                  75                  80
Leu His Asp Val Val Glu Val Ala Gly Leu Thr Ser Phe Ser Phe Gly
                85                  90                  95
Glu Asp Asp Asp Cys Arg Tyr Val Met Ile Phe Lys Lys Glu Phe Ala
            100                 105                 110
Pro Ser Asp Glu Glu Leu Asp Ser Tyr Arg Arg Gly Glu Glu Trp Asp
        115                 120                 125
Pro Gln Lys Ala Glu Glu Lys Arg Lys Leu Lys Glu Leu Ala Gln Arg
    130                 135                 140
Gln Glu Glu Glu Ala Ala Gln Gln Gly Pro Val Val Val Ser Pro Ala
145                 150                 155                 160
Ser Asp Tyr Lys Asp Lys Tyr Ser His Leu Ile Gly Lys Gly Ala Ala
                165                 170                 175
Lys Asp Ala Ala His Met Leu Gln Ala Asn Lys Thr Tyr Gly Cys Val
            180                 185                 190
Pro Val Ala Asn Lys Arg Asp Thr Arg Ser Ile Glu Glu Ala Met Asn
        195                 200                 205
Glu Ile Arg Ala Lys Lys Arg Leu Arg Gln Ser Gly Glu Glu Leu Pro
    210                 215                 220
Pro Thr Ser Arg Arg Pro Ala Gln Leu Pro Leu Thr Pro Gly Ala Gly
225                 230                 235                 240
Gln Gly Ala Gly Arg Asp Lys Ala Ala Ala Ile Arg Ala His Pro Gly
                245                 250                 255
Ala Pro Pro Leu Asn His Leu Leu Pro Ala Val Pro Gln Ala Gly Gly
            260                 265                 270
Lys Gln Val Phe Asp Leu Ser Pro Leu Glu Leu Gly Tyr Val Arg Gly
        275                 280                 285
Met Cys Val Cys Val
    290
```

What is claimed is:

1. An isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity over its entire length to a nucleotide sequence encoding the CBCALD05 polypeptide of SEQ ID NO:2; or a nucleotide sequence complementary to said isolated polynucleotide.

2. The polynucleotide of claim 1 wherein said polynucleotide comprises the nucleotide sequence contained in SEQ ID NO:1 encoding the CBCALD05 polypeptide of SEQ ID NO2.

3. The polynucleotide of claim 1 wherein said polynucleotide comprises a nucleotide sequence that is at least 80% identical to that of SEQ ID NO: 1 over its entire length.

4. The polynucleotide of claim 3 which is polynucleotide of SEQ ID NO: 1.

5. The polynucleotide of claim 1 which is DNA or RNA.

6. A DNA or RNA molecule comprising an expression system, wherein said expression system is capable of producing a CBCALD05 polypeptide comprising an amino acid sequence, which has at least 80% identity with the polypeptide of SEQ ID NO:2 when said expression system is present in a compatible host cell.

7. A host cell comprising the expression system of claim 6.

8. A process for producing a CBCALD05 polypeptide comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture.

9. A process for producing a cell which produces a CBCALD05 polypeptide thereof comprising transforming or transfecting a host cell with the expression system of claim 6 such that the host cell, under appropriate culture conditions, produces a CBCALD05 polypeptide.

10. A CBCALD05 polypeptide comprising an amino acid sequence which is at least 80% identical to the amino acid sequence of SEQ ID NO:2 over its entire length.

11. The polypeptide of claim 10 which comprises the amino acid sequence of SEQ ID NO:2.

12. An antibody immunospecific for the CBCALD05 polypeptide of claim 10.

13. A method fc the treatment of a subject in need of enhanced activity or expression of CBCALD05 polypeptide of claim 10 comprising:

(a) administering to the subject a therapeutically effective amount of an agonist to said polypeptide; and/or

(b) providing to the subject an isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity to a nucleotide sequence encoding the CBCALD05 polypeptide of SEQ ID NO:2 over its entire length; or a nucleotide sequence complementary to said nucleotide sequence in a form so as to effect production of said polypeptide activity *in vivo*.

14. A method for the treatment of a subject having need to inhibit activity or expression of CBCALD05 polypeptide of claim 10 comprising:

(a) administering to the subject a therapeutically effective amount of an antagonist to said polypeptide; and/or

(b) administering to the subject a nucleic acid molecule that inhibits the expression of the nucleotide sequence encoding said polypeptide; and/or

(c) administering to the subject a therapeutically effective amount of a polypeptide that competes with said polypeptide for its ligand, substrate , or receptor.

15. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of CBCALD05 polypeptide of claim 10 in a subject comprising:

(a) determining the presence or absence of a mutation in the nucleotide sequence encoding said CBCALD05 polypeptide in the genome of said subject; and/or

(b) analyzing for the presence or amount of the CBCALD05 polypeptide expression in a sample derived from said subject.

16. A method for identifying compounds which inhibit (antagonize) or agonize the CBCALD05 polypeptide of claim 10 which comprises:

(a) contacting a candidate compound with cells which express the CBCALD05 polypeptide (or cell membrane expressing CBCALD05 polypeptide) or respond to CBCALD05 polypeptide; and

(b) observing the binding, or stimulation or inhibition of a functional response; or comparing the ability of the cells (or cell membrane) which were contacted with the candidate compounds with the same cells which were not contacted for CBCALD05 polypeptide activity.

17. An agonist identified by the method of claim 16.

18. An antagonist identified by the method of claim 16.

19. A recombinant host cell produced by a method of Claim 9 or a membrane thereof expressing a CBCALD05 polypeptide.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN98/00010

A. CLASSIFICATION OF SUBJECT MATTER

IPC[6] C12N15/12, C07K14/47, C12N15/10

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched(classification system followed by classification symbols)

IPC[6] C12N15/12, C07K14/47, C12N 15/10

Documentation searched other than minimum documentation to the extent that such documents are included in the field searched

Chinese Patnets, Chinese Scientific and Technical Journals

Electronic data base consulted during the international search(name of data base and, where practicable, search terms used)

GenBank, EMBL, DDBJ, PDB, SwissProt, SPupdate, PIR, WPI

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant claim No.
A	Reprod. Fertil. Devo 6, 761-770 (1994) Beaton S. et al. ' Cloning and characterisation of a fox sperm protein FSA-1 ' see abstract	1-6 10-11
A	Nature Genet. 6 (3), 236-244 (1994) Cross, S. H. et al. ' Purification of CpG islands using a methylated DNA binding column ' see abstract	1-6 10-11

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason(as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 August 1998 (15. 08. 98)	10 SEP 1998 (10.09.98)

Name and mailing address of the ISA/	Authorized officer Xie Shunxing
The Chinese Patent Office 6, Xitucheng Road, Haidian District, Beijing, 100088, China	Telephone No. 86-010-62093101
Facsimile No. 86-010-62019451	

谢顺星 印

Form PCT/ISA/210(second sheet)(July 1992)

European Patent Office

PARTIAL EUROPEAN SEARCH REPORT

under Rule 46, paragraph 1 of the European Patent Convention

Application Number EP 04 00 1196

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	WO 02/18420 A (LION BIOSCIENCE AG; ALBERS, MICHAEL; ELLWANGER, SILVIA; KOEGL, MANFRED) 7 March 2002 (2002-03-07) * claims 1,12; sequences 28-30 *	1-14	C07K14/47 C07K14/705 C07K16/18 G01N33/53 C12Q1/68
X	WO 99/36521 A (SHANGHAI SECOND MEDICAL UNIVERSITY; MAO, MAO; FU, GANG; ZHANG, QING-HU) 22 July 1999 (1999-07-22) * page 20 - page 21; claims 4,11; sequences 1,2 *	1-14	C12P21/00 C12N15/11 C12N15/12 C12N9/00 G06F19/00 C07K14/47
P,X	DATABASE EMBL 9 September 2003 (2003-09-09), ISOGAI T., YAMAMOTO J.;: "Homo sapiens cDNA FLJ41070 fis, clone 3NB692002685, highly similar to Homo sapiens sperm acrosomal protein mRNA." XP002332133 Database accession no. AK123065 * the whole document *	1-14	
D,A	EP 1 195 434 A (TAISHO PHARMACEUTICAL CO., LTD; HELIX RESEARCH INSTITUTE) 10 April 2002 (2002-04-10) * the whole document *	1-14	

TECHNICAL FIELDS SEARCHED (Int.Cl.7)

C07K

LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to severalinventions or groups of inventions, namely:

see sheet B

The present partial European search report has been drawn up forthose parts of the European patent application which relate to the invention first mentioned in the claims.

Place of search	Date of completion of the search	Examiner
Munich	16 June 2005	Dumont, E

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C27)